Journal of Global Change Data & Discovery. 2021, 5(2): 175-180
www.geodoi.ac.cn

Global Change Research Data Publishing & Repository

Monthly Relative Humidity Dataset with 1-km Resolution in Nine Provinces in the Yellow River Basin (2000–2015)

Cai, H. Y.[1*] Jiang, X.[1,2] Yang, X. H.[1,2]

1. State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: Relative humidity is closely related to the diffusion, migration, and transformation of air pollutants and is an important index for regional environmental quality assessment. Based on the observation data from 940 meteorological stations in nine provinces and their surrounding areas in the Yellow River Basin, we generated the relative humidity dataset with 1-km resolution for nine provinces in the Yellow River Basin (2000–2015), using the ANUSPLIN software platform. The cubic spline function was applied for interpolation, with elevation as an independent covariate. The temporal resolution of the data was month, the spatial resolution was 1 km, and the projection was based on Albers Conical Equal Area with the coordinate system of WGS-84. The data are provided in the .tif format.

Keywords: Yellow River Basin; relative humidity; 2000–2015

DOI: https://doi.org/10.3974/geodp.2021.02.09

CSTR: https://cstr.escience.org.cn/CSTR:20146.14.2021.02.09

Dataset Availability Statement:

The dataset supporting this paper was published and is accessible through the *Digital Journal of Global Change Data Repository* at: https://doi.org/10.3974/geodb.2021.03.12.V1 or https://cstr.escience.org.cn/CSTR:20146.11.2021.03.12.V1.

1 Introduction

Relative humidity is jointly determined by the water vapor content in the atmosphere and the air temperature; it is an important index to measure the degree of regional dryness and wetness[1–3]. These two parameters are directly related to the diffusion, migration, and transformation of pollutants in the atmosphere[4–6] and closely related to human health and comfort. In this sense, relative humidity is also an important indicator for the evaluation of regional environmental quality and urban livability.

The Yellow River Basin spans over three major regions in China, covering nine provinces.

Received: 09-02-2021; **Accepted:** 20-04-2021; **Published:** 25-06-2021

Foundations: Ministry of Science and Technology of P. R. China (2017YFC0503803)

***Corresponding Author:** Cai, H.Y. Y-8555-2019, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, caihy@igsnrr.ac.cn

Data Citation: [1] Cai, H. Y., Jiang, X., Yang, X. H. Monthly relative humidity dataset with 1-km resolution in nine provinces in the Yellow River Basin (2000–2015) [J]. *Journal of Global Change Data & Discovery*, 2021, 5(2): 175–180. https://doi.org/10.3974/geodp.2021.02.09. https://cstr.escience.org.cn/CSTR:20146.14.2021.02.09.

[2] Cai, H. Y., Jiang, X., Yang, X. H. 1 km-monthly humidity dataset in Yellow River Basin covering nine provinces of China (2000–2015) [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2021. https://doi.org/10.3974/geodb.2021.03.12.V1. https://cstr.escience.org.cn/CSTR:20146.11.2021.03.12.V1.

It is a key area of ecological protection and economic development, leading to significant ecological issues. The Yellow River Basin is one of the typical climate-sensitive areas in China, most of which are located in arid and semi-arid areas, with an uneven distribution of water resources and a fragile ecological environment. In the last 20 years, due to the unbalanced use of water resources, the lake area has shrunk, and floods and droughts have occurred frequently[7,8]. The high level of industrialization in the area has led to serious air pollution[9]. In September 2019, Chairman Xi Jinping stated that the ecological protection of the Yellow River Basin should be upgraded to a major national strategy[10].

In this context, temporal and spatial distribution data of relative humidity in the Yellow River basin can help to deepen our understanding of the dry and wet changes in this region, provide scientific support for regional ecological environment quality assessment and management, and have important significance for implementing ecological protection strategies.

2 Metadata of the Dataset

The metadata of the 1 km-monthly humidity dataset in Yellow River Basin covering nine provinces of China (2000–2015)[11] is summarized in Table 1. It includes the dataset full name, short name, authors, year of the dataset, temporal resolution, spatial resolution, data format, data size, data files, data publisher, and data sharing policy, etc.

Table 1 Metadata summary of the 1 km-monthly humidity dataset in Yellow River Basin covering nine provinces of China (2000–2015)

Items	Description
Dataset full name	1 km-monthly humidity dataset in Yellow River Basin covering nine provinces of China (2000–2015)
Dataset short name	RHU_9PYRB
Authors	Cai, H.Y. Y-8555-2019, State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, caihy@igsnrr.ac.cn Jiang, X. AAE-1541-2021, State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences; University of Chinese Academy of Sciences, jiangx.20b@igsnrr.ac.cn Yang, X. H. AAC-8887-2021, State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences; University of Chinese Academy of Sciences, yangxh@lreis.ac.cn
Geographical region	Nine provinces
Year	2000–2015
Temporal resolution	Monthly
Spatial resolution	1 km
Data format	.tif
Data size	1.45 GB in compression
Data files	(1) Folder ‘Studyarea’ is the boundary data in .shp format, (2) Folder ‘9PYRB_RHU’ includes monthly spatial and temporal distribution data of relative humidity from 2000 to 2015 in .tif format. The Data name contains its phase information, such as “9PYRB_ Rhu20001.tif ” is the relative humidity data of nine provinces in the Yellow River Basin in January 2000
Foundation	Ministry of Science and Technology of P. R. China (2017YFC0503803)
Data publisher	Global Change Research Data Publishing & Repository, http://www.geodoi.ac.cn
Address	No. 11A, Datun Road, Chaoyang District, Beijing 100101, China
Data sharing policy	***Data*** from the Global Change Research Data Publishing & Repository includes metadata, datasets (in the *Digital Journal of Global Change Data Repository*), and publications (in the *Journal of Global Change Data & Discovery*). ***Data*** sharing policy includes: (1) ***Data*** are openly available and can be free downloaded via the Internet; (2) End users are encouraged to use ***Data*** subject to citation; (3) Users, who are by definition also value-added service providers, are welcome to redistribute ***Data*** subject to written permission from the GCdataPR Editorial Office and the issuance of a ***Data*** redistribution license; and (4) If ***Data*** are used to compile new datasets, the ‘ten per cent principal’ should be followed such that ***Data*** records utilized should not surpass 10% of the new dataset contents, while sources should be clearly noted in suitable places in the new dataset[12]
Communication and searchable system	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS/ISC, GEOSS

3 Methods

3.1 Algorithm Principle

The ANUSPLIN package is a tool based on ordinary-thin disk and local-thin disk spline function for multivariate interpolation and suitable for interpolation of time-series meteorological data. Local thin-disk smooth spline is an extension of the prototype of thin-disk smooth spline, which allows the introduction of a covariant quantum model in addition to ordinary spline-independent variables. The output statistics are best interpreted in relation to the local-thin disk spline model for N observed data values z_i given by:

$$z_i = f(x_i) + b^T y_i + e_i (i = 1, 2, \ldots, N) \tag{1}$$

where x_i is a d-dimensional vector of spline independent variable; f is an unknown smooth function of the x_i; y_i is a p-dimensional vector of independent covariates; b is an unknown p-dimensional vector of coefficient of y_i; e_i is the random error[13].

The function f and coefficient b are determined by Equation (2), representing the least square method:

$$\sum\nolimits_i^n \left[\frac{z_i - f(x_i) - b^T y_i}{w_i} \right]^2 + \rho J_m(f) \tag{2}$$

where $J_m(f)$ is a measure of the complexity of function $f(x_i)$, the "roughness penalty" defined in terms of mth order partial derivatives of f and ρ is a smooth parameter, balancing data fidelity and surface roughness; w_i is termed the known relative error.

In ANUSPLIN interpolation, the optimal model is judged by generalised cross validation (GCV), generalised maximum likelihood (GML), mean square error (MSE), and signal. When GCV is smaller and signal is less than half of the number of sites, it is assumed that the best smoothing parameter has been found in the fitting process. The selected model is suitable for interpolation, and the effect of surface fitting is better.

3.2 Technical Route

The main process of dataset production included the following (Figure 1):

(1) Data preprocessing of the relative humidity records from meteorological stations. First, strict quality inspection was carried out on the original relative humidity data (. txt format), with monthly measurements from 2000 to 2015; the data were obtained from the National Meteorological Information Center of China. In general, the quality of the observation data in the nine provinces of the Yellow River Basin was good, and the data were relatively complete. Here, we detected and processed data outliers and eliminated the meteorological stations without measurements. The preprocessed data were stored in ASCII format to prepare for interpolation;

(2) Optimization of the interpolation scheme. Four groups of interpolation schemes were designed, as shown in Table 2, and the schemes were tested with the relative humidity in 2015. According to the GCV value, the interpolation scheme was best when the independent variables were longitude, latitude, and elevation, the independent covariate was elevation, and the number of splines was three;

(3) Data interpolation. The monthly relative humidity records from meteorological stations during 2000 to 2015 was interpolated by the optimal scheme;

(4) Accuracy verification of simulation results. The deviation and relative error index values were used to verify the results by comparing the observed values with the simulated ones.

(5) The monthly relative humidity dataset with 1-km resolution in nine provinces in the Yellow River Basin from 2000 to 2015 was generated.

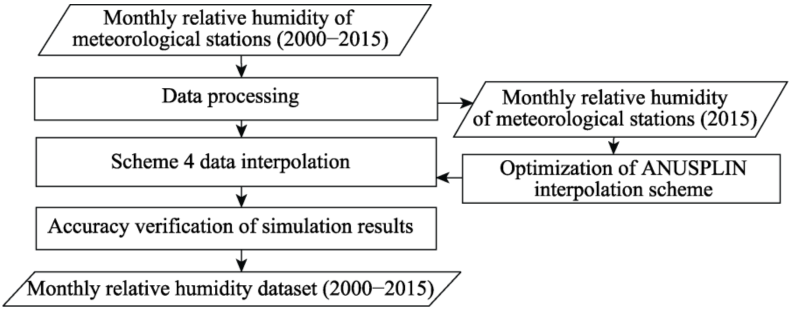

Figure 1 Flow chart of the data processing

Table 2 ANUSPLIN interpolation model schemes and GCVs

Number	Independent variable	Independent covariates	Spline number	GCV
1	Longitude, latitude	–	2	13.5
2	Longitude, latitude, elevation	Elevation	2	13.4
3	Longitude, latitude	Elevation	3	12.2
4	Longitude, latitude, elevation	Elevation	3	12.1

4 Data Results and Validation

4.1 Data Composition

The dataset includes two folders, one is for the boundary data, and the other is for monthly spatial and temporal distribution data of relative humidity.

4.2 Spatial Distribution Data

Taking March, June, September, and December 2015 as examples, the spatial distribution characteristics of the relative humidity in the nine provinces are shown in Figure 2. The relative humidity in summer and autumn (June and September) was higher than that in spring and winter (March and December); in winter, humidity was slightly higher than that in spring. The variation of the relative humidity in different months was closely related to the regional climate conditions, positively related to precipitation, and negatively related to temperature[14]. Regarding spatial distribution, the relative humidity in the southeast was higher than that in the northwest in these 4 months, which is consistent with the spatial distribution of rainfall in the basin[15].

4.3 Data Validation

Based on the measured relative humidity of the nine provinces and surrounding meteorological stations in the Yellow River Basin from January to December 2015, as well as on interpolation results, the deviation *MBE* and relative error *Er* were selected for verification, as follows:

$$MBE = \frac{1}{N}\sum_{i=1}^{N}(P_i - O_i) \tag{3}$$

$$Er = \frac{P_i - O_i}{O_i} \times 100\% \tag{4}$$

where P_i is the interpolation result and O_i is the observation data of meteorological stations. The closer $|MBE|$ and $|Er|$ are to 0, the closer the interpolation is to the observed value, and the more reliable the interpolation results are.

As shown in Figure 3, taking the year of verification as an example, the $|MBE|$ value is between 0.00 and 0.10 and the $|Er|$ value between 3.6% and 5.6%, indicating that the interpolation result accuracy is high.

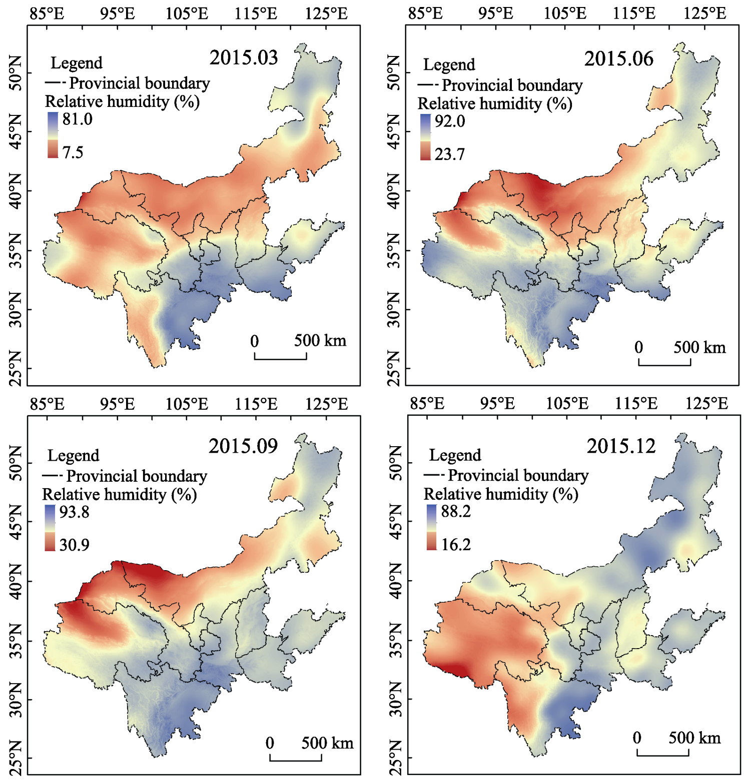

Figure 2 Spatial distribution of relative humidity in nine provinces of the Yellow River Basin in different months of 2015

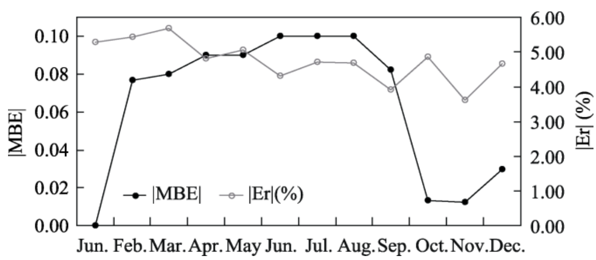

Figure 3 Validation of interpolation accuracy of relative humidity in nine provinces of the Yellow River Basin in 2015

5 Discussion and Conclusion

Based on the relative humidity records from meteorological stations in nine provinces and surrounding areas in the Yellow River Basin, this study generated monthly relative humidity dataset with 1-km resolution from 2000 to 2015. The cubic spline function was selected, with elevation as the independent covariate for spatial interpolation. According to the accuracy verification, the interpolation results are reliable. This dataset revealed the spatial and temporal characteristics of relative humidity changes in nine provinces in the Yellow River Basin from 2000 to 2015, providing reliable data for the assessment and management of the ecological environment in the region.

Our results showed that the relative humidity in summer and autumn was higher than that in winter and spring, which was consistent with the temporal pattern of regional rainfall. In terms of spatial distribution, the relative humidity of the entire region was high in the south and low in the north, as well as high in the east and low in the west. In this study, four interpolation schemes were tested for the data interpolation, and the scheme with elevation, longitude, and latitude as independent variables and elevation as independent covariate delivered the best results. However, it should be noted that the optimal scheme may vary with different meteorological indicators or regions. In addition, the dataset has the disadvantage that when elevation is used as an independent covariate, the interpolation range is limited by the elevation range. Because of the lack of elevation data in the northern coastal area of Shandong province, this dataset was missing in the corresponding area. Nevertheless, this is negligible and does not affect the use of the whole-area data.

***Author* Contributions**

Cai, H. Y. carried out the overall design, data analysis, and paper writing for dataset production; Jiang, X. is responsible for the paper writing as well as the processing of relative humidity data; Yang, X. H. is responsible for the paper improvement.

Conflicts of Interest

The authors report no conflicts of interest.

References

[1] Liu, M. C., Yang, X. L., Yin, Y. C., *et al*. Climate characteristics and forecast of relative humidity in Wuwei city [J]. *Arid Zone Research*, 2012, 29(4): 654−659.

[2] Du, J., Cooper, F., Fueglistaler, S. Statistical analysis of global variations of atmospheric relative humidity as observed by AIRS [J]. *Journal of Geophysical Research Atmospheres*, 2012, 117: D12315.

[3] Liu, Y. Y., Li, Y. F., Xie, J. F., *et al*. Climate change characteristics of free atmospheric humidity and its relationship with temperature and precipitation in Northeast China [J]. *Scientia Geographica Sinica*, 2016, 36(4): 628−636.

[4] Eck, T. F., Holben, B. N., Kim, J., *et al*. Influence of cloud, fog, and high relative humidity during pollution transport events in South Korea: aerosol properties and $PM_{2.5}$ variability [J]. *Atmospheric Environment*, 2020, 232: 1−16.

[5] Zhang, Z., Wang, W. H., Wang, H. M., *et al*. Effects on humidity on absorbing aerosol index [J]. *Journal of Remote Sensing*, 2019, 23(6): 1177−1185.

[6] Suonan, K. Z., Ren, G. Y., Jia, W. Q., *et al*. Climatological characteristics and long-term trend of relative humidity in Wuhan [J]. *Climatic and Environmental Research*, 2018, 23(6): 715−724.

[7] Huang, J. P., Zhang, G. L., Yu, H. P., *et al*. Characteristic of climate change in the Yellow River basin during recent 40 years [J]. *Journal of Hydraulic Engineering*, 2020, 51(9): 1048−1058.

[8] Li, Y. Y. Research on the drought assessment-propagation-driving-prediction under the climate and land use land cover change scenarios [D]. Xi'An: Xi'an University of Technology, 2018.

[9] Wu, D. Hazy weather research in China in the last decade: a review [J]. *Acta Scientiae Circumstantiae*, 2012, 32(2): 257−269.

[10] Xi, J. P. Speech at the symposium on ecological protection and high-quality development of the Yellow River Basin [J]. *Qiushi*, 2019, 20: 1−5.

[11] Cai, H. Y., Jiang, X., Yang, X. H. 1 km-monthly humidity dataset in Yellow River Basin covering nine provinces of China (2000–2015) [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2021. https://doi.org/10.3974/geodb.2021.03.12.V1. https://cstr.escience.org.cn/CSTR:20146.11.2021.03.12.V1.

[12] GCdataPR Editorial Office. GCdataPR data sharing policy [OL]. https://doi.org/10.3974/dp.policy.2014.05 (Update 2017).

[13] Liu, Z. H., Li, L. T., McVicar, T. R., *et al*. Introduction of professional interpolation software for meteorology data: ANUSPLIN [J]. *Meteorological Monthly*, 2008, 34(2): 92−100.

[14] Xu, R. L., Li, B. F., Lian, L. S. Quantitative relationship between the spatiotemporal change of relative humidity and climatic factors in the arid region of Northwest China from 1960 to 2015 [J]. *Research of Soil and Water Conservation*, 2020, 27(6): 233−239, 246.

[15] Shao, X. M., Yan, C. R., Wei H. B. Spatial and temporal structure of precipitation in the Yellow River Basin based on Kriging method [J]. *Chinese Journal of Agrometeorology*, 2006, 27(2): 65−69.